Virology 434 (2012) 202–209

Contents lists available at SciVerse ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Review

Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information

Peter C. Fineran [a,*], Emmanuelle Charpentier [b,**]

[a] Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[b] The Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Department of Molecular Biology, Umeå University, Umeå S-90187, Sweden

ARTICLE INFO

Available online 2 November 2012

Keywords:
CRISPR-Cas
Cas1
Cas2
Bacteria
Archaea
Viruses
Mobile genetic elements
Horizontal gene transfer
Adaptive immunity
Spacer acquisition

ABSTRACT

Multiple organisms face the threat of viral infections. To combat phage invasion, bacteria and archaea have evolved an adaptive mechanism of protection against exogenic mobile genetic elements, called CRISPR-Cas. In this defense strategy, phage infection is memorized via acquisition of a short invader sequence, called a spacer, into the CRISPR locus of the host genome. Upon repeated infection, the 'vaccinated' host expresses the spacer as a precursor RNA, which is processed into a mature CRISPR RNA (crRNA) that guides an endonuclease to the matching invader for its ultimate destruction. Recent efforts have uncovered molecular details underlying the crRNA biogenesis and interference steps. However, until recently the step of adaptation had remained largely uninvestigated. In this minireview, we focus on recent publications that have begun to reveal molecular insights into the adaptive step of CRISPR-Cas immunity, which is required for the development of the heritable memory of the host against viruses.

Contents

Introduction

During their lifetime, bacteria and archaea face the constant threat of invading foreign DNA, mainly mobile genetic elements such as phages, plasmids, transposons and genomic islands. A gain of novel genetic traits can have a beneficial or detrimental consequence on the host. For example, the horizontal transfer of genetic elements contributes largely to the acquisition of antibiotic resistance by environmental and clinical bacteria. In addition, virulence determinants can be acquired, leading to toxigenic conversion of bacterial strains. A particular threat to bacteria and archaea are their viral predators. The global phage population is genetically diverse, their abundance exceeds bacterial numbers by an order of magnitude and an estimated 10^{25} infections occur every second (Hendrix, 2003; Weinbauer, 2004; Wommack and Colwell, 2000). Therefore, an arms race is said to exist between prokaryotes and their viruses and to survive phage infection, and control the flow of genetic information, bacteria and archaea have evolved diverse defense strategies (Labrie et al., 2010).

To counteract viral infections, eukaryotic organisms launch an immune response consisting of innate (or non-specific) and adaptive (or specific) mechanisms. Most viral infections are halted by the first line of innate immune defenses that are continuously

* Corresponding author.
** Corresponding author.
E-mail addresses: peter.fineran@otago.ac.nz (P.C. Fineran), emmanuelle.charpentier@mims.umu.se (E. Charpentier).

0042-6822/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2012.10.003

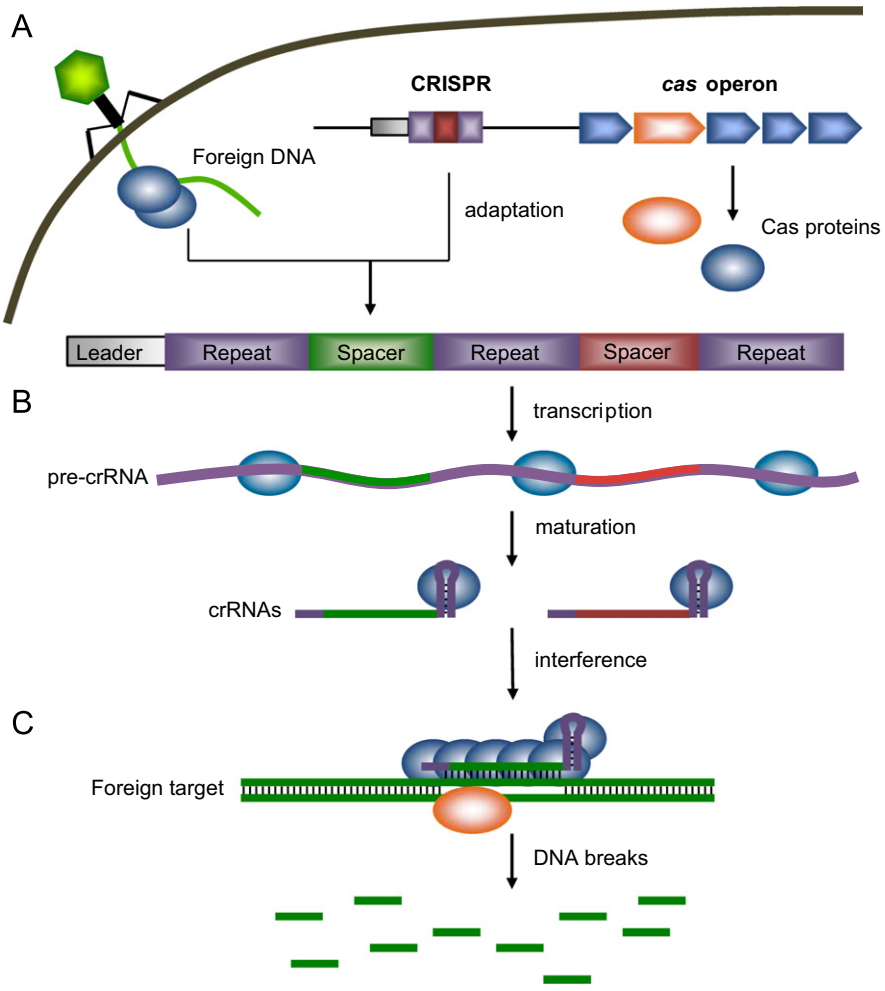

Fig. 1. Overview of CRISPR-Cas immunity to viruses of bacteria and archaea. (A) In adaptation, phage infection is recognized by Cas proteins (presumably the core Cas1 and Cas2) and a short sequence of the phage DNA (termed a protospacer, proposed to be now termed precursor-spacer or pre-spacer (Westra and Brouns, 2012)) is added to the leader end of the CRISPR array, resulting in a new spacer sequence and a duplicated repeat. Represented are the *cas* operon encoding the Cas proteins and the closely associated CRISPR array, composed of the leader sequence followed by a series of repeats-spacer units. (B) Transcription of the CRISPR array from a promoter within the leader sequence results in a precursor CRISPR RNA (pre-crRNA) transcript. The pre-crRNA is matured into individual crRNAs by a process involving Cas proteins. (C) The mature crRNAs form a ribonucleoprotein complex, which targets nucleic acids that are complementary to the spacer sequence in the crRNAs. In some cases a separate Cas nuclease (orange) is recruited, resulting in interference and destruction of the nucleic acid target. The general schematic is based on the type I system, and differences exist between the type I, II and III. For details, see the text.

active in the host without exposure to any virus. In cases when viral replication outpaces innate defenses, the host then mounts the adaptive response. Similar defense strategies against viral infection apply to microorganisms like bacteria and archaea (Bikard and Marraffini, 2012). Innate immunity against phages can be considered to involve the mechanisms of abortive infection, mutation of host receptors or restriction/modification of the incoming foreign DNA. However, in most cases these systems are not truly innate since they also display a degree of specificity. Recently, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) has been discovered as an adaptive defense mechanism against phages (reviewed recently by Bhaya et al. (2011); Deveau et al. (2010); Horvath and Barrangou (2010); Marraffini and Sontheimer (2010); Terns and Terns (2011); van der Oost et al. (2009); Wiedenheft et al. (2012)). The system is heritable, widespread among bacteria and archaea and active in immunity against various mobile genetic elements.

CRISPR-Cas immunity is mediated by RNA and protein components that function together in ribonucleoprotein complexes. The CRISPR-Cas immune strategy consists of an adaptive phase with acquisition of memory, a biogenesis phase to generate the guide RNA components and a phase of interference of the invading cognate nucleic acids by ribonucleoprotein complexes consisting of Cas proteins and the guide RNAs (Fig. 1) (Bhaya et al., 2011; Deveau et al., 2010; Horvath and Barrangou, 2010; Marraffini and Sontheimer, 2010; Terns and Terns, 2011; van der Oost et al., 2009; Wiedenheft et al., 2012). "Adaptive" refers here to the specificity of the immune response that is customized to a particular foreign invader. A key feature in the adaptation phase of CRISPR-Cas is memory, whereby a repeated infection by the same phage is stopped immediately by the specific response. The loci are commonly composed of an array of repeat-spacer sequences encoding the RNA components and an operon of *cas* genes encoding the protein components. The array consists of a leader sequence followed by a succession of short identical repeats regularly interspaced by short spacer sequences. The spacer sequences originate from previous encounters with foreign genetic material and thus function as a memory bank that will recognize the same genetic encounter upon a repeated infection. Briefly, CRISPR-Cas immunity operates as follows. Upon infection with the genetic intruder, a short sequence of the invading DNA (termed a protospacer, proposed to be now termed precursor-spacer

or pre-spacer (Westra and Brouns, 2012)) is inserted in the CRISPR array as a spacer sequence. The repeat-spacer array is transcribed from a promoter region present in the leader sequence as a precursor CRISPR RNA (pre-crRNA) molecule. The pre-crRNA undergoes one or more maturation events to generate the individual mature crRNAs that are composed each of repeat portion(s) and a targeting spacer portion. The mature crRNAs then function as guide RNAs that direct the Cas protein(s), in a sequence-specific manner, to cleave the invading nucleic acids.

The first description of CRISPR elements dates back to 1987 when Ishino et al. (1987) discovered a series of short palindromic sequences regularly repeated and separated by short unique sequences on the genome of *Escherichia coli*. Later, CRISPR arrays were also detected in archaea (Groenen et al., 1993; Masepohl et al., 1996; Mojica et al., 2005) and three independent studies identified the viral and plasmid source of spacer sequences (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). The observation of spacers matching foreign genetic elements, combined with a detailed bioinformatic analysis of the Cas proteins revealing putative nuclease and helicase domains, led to the proposal that CRISPR-Cas functions as an RNA-mediated adaptive immune system (Makarova et al., 2006). In 2007, CRISPR-Cas activity in adaptive immunity against phages was demonstrated for the first time in the laboratory with infection experiments of the Gram-positive lactic acid bacterium *Streptococcus thermophilus* by lytic phages (Barrangou et al., 2007). The study raised considerable attention in the scientific community, which recognized the potential to use the adaptive feature of CRISPR-Cas immunity as a new tool to control phage infection in the dairy industry (Barrangou and Horvath, 2012). Since 2007, there has also been an exponential interest in understanding how the immunity functions at the molecular level. Considerable efforts from a community of microbiologists and structural biologists have led to significant novel findings in the mechanisms of crRNA biogenesis (Brouns et al., 2008; Carte et al., 2008; Deltcheva et al., 2011; Hatoum-Aslan et al., 2011; Haurwitz et al., 2010; Nam et al., 2012b; Przybilski et al., 2011) and interference with nucleic acids (Garneau et al., 2010; Jinek et al., 2012; Westra et al., 2012).

The CRISPR-Cas systems have undergone rapid evolution and a recent classification of the systems led to a distribution into three types (*i.e.* I, II and III) further sub-grouped into several sub-types (*i.e.* I-A to I-F, II-A and II-B, III-A and III-B) characterized by Cas protein signatures (*e.g.* Cas3, Cas9, Cas10 for types I, II and III). Although the three systems share common principles in the immunity steps, they utilize distinct molecular mechanisms for crRNA biogenesis and interference. For example, in the bacterial and archaeal systems I and III, Cas6-like proteins are endoribonucleases that cleave pre-crRNA (Brouns et al., 2008; Carte et al., 2008; Ebihara et al., 2006; Hatoum-Aslan et al., 2011; Haurwitz et al., 2010; Przybilski et al., 2011; Richter et al., 2012b), while in the bacterial type II system, processing of the precursor molecule requires the concerted action of a small *trans*-activating RNA (tracrRNA), the protein Cas9 (formerly named Csn1) and the non-Cas endoribonuclease III (RNase III) acting as a dicing effector (Deltcheva et al., 2011). Furthermore, the mature crRNAs in type I and III systems guide a complex of Cas proteins, referred to as the Cascade-like complexes, to the cognate nucleic acids (invading protospacer) for subsequent cleavage by an effector Cas endonuclease (Beloglazova et al., 2011; Howard et al., 2011; Jore et al., 2011; Lintner et al., 2011; Mulepati and Bailey, 2011; Sinkunas et al., 2011; Westra et al., 2012; Wiedenheft et al., 2011a; Wiedenheft et al., 2011b). In system II, no complex of Cas proteins is implicated. Instead, a dual-RNA structure formed between the mature forms of tracrRNA and crRNA guides the single protein Cas9 to cleave the DNA (Jinek et al., 2012).

Despite major advances in our understanding of the crRNA biogenesis and interference steps, until recently there has been little insight into the acquisition of immunity, whereby new spacers are added to the CRISPR arrays. A number of critical studies have begun to elucidate details of the acquisition process, which will be the focus of this minireview.

How is new information acquired?

2012 has seen major advances in our understanding of spacer acquisition, whereby the repeat is duplicated and a new spacer is added at the leader end of the array. The first key experimental demonstration of spacer incorporation into CRISPR arrays was in 2007 in the type II-A system of *Streptococcus thermophilus* (Barrangou et al., 2007). Barrangou et al. challenged *S. thermophilus* with two phages, either separately or in combination, and selected for phage-resistant survivors. The CRISPR arrays in these strains had acquired between one and four spacers from the invading phages at the leader end of the array. The two bacteriophages were unique, yet shared regions of sequence identity and cross-resistance against both phages was observed when one spacer had sequence similarity to protospacers in both phages. Furthermore, spacers were incorporated from either the sense or antisense strand of the phage genome and there seemed to be no preference for which part of the phage genome was selected for integration in the arrays. The same group identified a short sequence adjacent to the protospacer target (Deveau et al., 2008), later termed the protospacer adjacent motif (PAM), that was required for interference and possibly acquisition. Later work in *S. thermophilus* provided similar evidence that spacers could be acquired from plasmids carrying antibiotic resistance genes (Garneau et al., 2010).

Many bioinformatic and metagenomic studies have also provided further evidence of the process of adaptation (Andersson and Banfield, 2008; Horvath et al., 2009; Horvath et al., 2008; Pride et al., 2012; Pride et al., 2011; Tyson and Banfield, 2008). These studies indicated that new spacers are added at the leader end of the array. However, this dogma was recently challenged when acquisition of new spacers internal to a CRISPR array was observed in wet laboratory experiments in *Sulfolobus solfataricus* (Erdmann and Garrett, 2012). Bioinformatic analyses have shown that CRISPR arrays are rapidly evolving in many bacteria. On this basis, the CRISPR-Cas loci can be used as a powerful phylogenetic marker to study the evolution of closely related strains over short time scales, as exemplified in *Yersinia pestis* (Drevet and Pourcel, 2012; Grissa et al., 2008; Grissa et al., 2007a; Grissa et al., 2007b; Vergnaud et al., 2007). There are exceptions however; the CRISPRs of *E. coli* have evolved very slowly and cannot be used effectively for studying recent evolutionary histories (Touchon et al., 2011; Touchon and Rocha, 2010). Strong evidence for the rapid ability of CRISPR-Cas systems to respond to viral infection and acquire spacers comes from metagenomic analyses of the population dynamics of mobile genetic elements, such as viruses, and bacteria in both environmental and human niches (Andersson and Banfield, 2008; Pride et al., 2012; Pride et al., 2011; Tyson and Banfield, 2008). However, with the exception of *S. thermophilus* (Barrangou et al., 2007; Deveau et al., 2008), spacer addition in other bacteria or in archaea was not reported in the laboratory until 2012. Spacer acquisition has now been detected under laboratory conditions in the *E. coli* type I-E (Datsenko et al., 2012; Swarts et al., 2012; Yosef et al., 2012), *Pseudomonas aeruginosa* type I-F (Cady et al., 2012), *Streptococcus agalactiae* type II-A (Lopez-Sanchez et al., 2012) and *S. solfataricus* type I and III-B (Erdmann and Garrett, 2012) systems. The studies of the *E. coli* type I-E system have provided the greatest insight into the

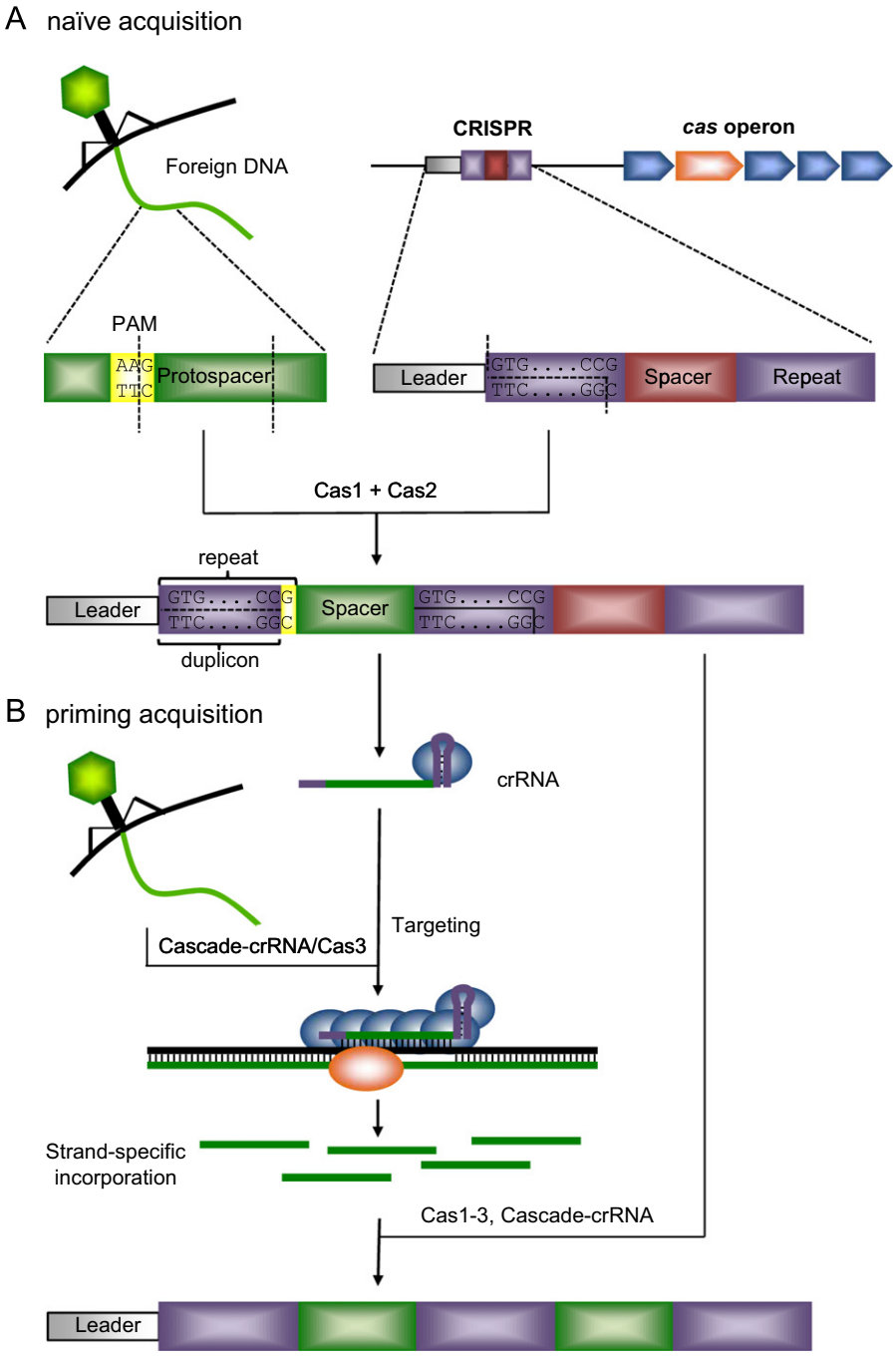

Fig. 2. Adaptation in the type I-E CRISPR-Cas system. (A) In naïve acquisition, infection with a phage that has not previously been encountered results in acquisition of a new spacer. The pre-spacer, including the final nt of the PAM, is acquired from the phage (dashed lines) and incorporated into the leader end of the array. The leader proximal repeat, with the exception of the final nt, is copied upon spacer incorporation (termed the duplicon) and the final nt of the PAM becomes the 3′ nt of the repeat (yellow). Naïve acquisition requires Cas1 and Cas2. (B) During priming acquisition, the presence of a targeting crRNA against the original phage, or an escape mutant phage, leads to incorporation of new spacers in a strand-specific manner. This process requires Cas1, Cas2, Cascade-crRNA and Cas3.

underlying mechanisms of spacer integration and will be the focus of the remainder of the review. The current model of this process is summarized in Fig. 2.

Which Cas proteins are required for spacer acquisition?

Cas1 and Cas2 are the only Cas proteins present in all subtypes of CRISPR-Cas systems (Haft et al., 2005; Makarova et al., 2006). Based on their conservation, tight genomic co-association and their predicted biochemical functions, Makarova et al. (2006) proposed that spacer integration was mediated by Cas1 in co-operation with Cas2. Overexpression of only Cas1 and Cas2 by Yosef et al. (2012) led to the incorporation of spacers derived from the expression plasmid and to a lesser extent, the chromosome. Both proteins were required since expression of either Cas1 or Cas2 alone did not stimulate spacer acquisition. In a complementary study by Datsenko et al. (2012), mutation of *cas1* or *cas2* abolished phage M13 spacer incorporation, whereas overexpression of only *cas1* and *cas2* promoted acquisition (Datsenko et al., 2012).

In both cases, Cascade and Cas3 were not essential for spacer integration. Thus, these studies show that Cas1 and Cas2 are both necessary for acquisition, which is consistent with their dispensable role in crRNA maturation and interference in type I-E (Brouns et al., 2008), type II-A (Deltcheva et al., 2011; Sapranauskas et al., 2011) and type III-A (Hatoum-Aslan et al., 2011) systems. These results are relevant to other CRISPR-Cas types because, as mentioned above, Cas1 and Cas2 are present in all systems (Makarova et al., 2006; Makarova et al., 2011b). However, in other systems additional proteins may be involved. For example, an insertion mutation of *csn2* inhibited spacer acquisition in the *S. thermophilus* type II-A system (Barrangou et al., 2007).

The genetic studies described above proved unequivocally that Cas1 and Cas2 are necessary for spacer acquisition, but they did not reveal the exact mechanistic roles of these proteins. A number of biochemical and structural investigations provide an indication of the functions of Cas1 (Babu et al., 2011; Han et al., 2009; Wiedenheft et al., 2009) and Cas2 (Beloglazova et al., 2008; Nam et al., 2012a; Samai et al., 2010). The first Cas1 protein to be studied biochemically was SSO1450 from *S. solfataricus*. The *S. solfataricus* Cas1 bound ss/dsDNA, ss/dsRNA and DNA–RNA hybrids with high-affinity (~18–50 nM). In addition, Cas1 bound substrates containing CRISPR repeat and spacer sequences, but with no increase in affinity compared with unrelated sequences, demonstrating no apparent sequence specificity. The ability to promote the hybridization of single-stranded DNA strands was also proposed (Han et al., 2009). In 2009, the crystal structure of the *P. aeruginosa* type I-F Cas1 was solved and was shown to be a Mn^{2+} or Mg^{2+}-dependent endonuclease, which cleaved dsDNA into short fragments of ~80 bp (Wiedenheft et al., 2009). The predicted size of spacer substrates for integration in this system was ~33 nt. Therefore, the ~80 bp size of these Cas1-derived fragments was longer than expected, suggesting that other proteins, such as Cas2, may be required to process the ~80 bp fragment to the correct size. The *Pseudomonas* Cas1 is a homodimer, which contains both a stirrup-like structure, approximately 20 Å in diameter, and a positively charged surface that jointly were proposed to be involved in DNA binding (Wiedenheft et al., 2012). The structure of the type I-E Cas1 from *E. coli* was subsequently solved and other structures are also available in the protein databank (PDB) (Babu et al., 2011). *E. coli* Cas1 cleaved ssRNA and ss/dsDNA substrates and also resolved Holliday junctions, which are branched DNA intermediates important for DNA integration and recombination events. The non-specific binding of the different Cas1 proteins in these studies is consistent with a role in spacer acquisition, whereby the candidate pre-spacers that are sampled are of different sequence. If Cas1 is involved in generation of spacer substrates from candidate pre-spacers, the only sequence specificity that might be required is likely to be the PAM (see later), but this has yet to be examined at the biochemical level.

The structures of several Cas2 proteins have also been solved and the proteins examined biochemically (Beloglazova et al., 2008; Nam et al., 2012a; Samai et al., 2010). Cas2 from *S. solfataricus* consists of a ferredoxin-like fold and cleaved ssRNA at U-rich sequences (Beloglazova et al., 2008). In contrast, when ssRNA or ssDNA binding or cleavage was assessed for *Desulfovibrio vulgaris* Cas2, no activity was observed (Samai et al., 2010). A recent study demonstrated that the *Bacillus halodurans* Cas2 dimer was a Mg^{2+}-dependent dsDNA-specific endonuclease, which generated ~120 bp fragments (Nam et al., 2012a). These authors analyzed the known Cas2 structures and proposed that a loop region required for discrimination between RNA and DNA substrates would provide an explanation for the different activities of the characterized Cas2 proteins. The strict requirement of both Cas1 and Cas2 for the integration event *in vivo* (Datsenko et al., 2012; Yosef et al., 2012), suggests that biochemical assays performed with proteins, candidate pre-spacer (with PAM) and leader-repeat substrates might be more revealing. We envisage that an integration-type protein/DNA complex mediates the adaptation stage.

One intriguing question about Cas1 and Cas2 is whether they function together directly via protein–protein interactions by analogy to the Cas complex formation that is required for type I and III interference. In *Thermoproteus tenax*, which contains a type I-A system, Cas1 and Cas2 are fused in a single protein, which interacts with both Cas4 and Csa1 (Plagens et al., 2012). These four genes constitute an operon and the complex has been termed Cascis (CRISPR associated complex for the integration of spacers). These results strengthen the possibility that Cas4 and Csa1 are involved in the acquisition phase (Makarova et al., 2011a; Makarova et al., 2011b; Plagens et al., 2012). In some CRISPR-Cas systems, Cas4 and Cas1 exist as a protein fusion, providing further evidence that they may function together in acquisition in the type I-A, I-B, I-C, I-D and II-B systems (Makarova et al., 2011a; Makarova et al., 2011b; van der Oost et al., 2009). The role of Cascis in adaptation is uncertain, but it is interesting that a Cas1–Cas3 complex is also formed, which might be involved in the adaptation stage of type I-F systems (Richter et al., 2012a). The Cas3 protein of the type I-F systems is a hybrid of Cas2–Cas3 and contains an N-terminal Cas2-like domain fused to Cas3 (Makarova et al., 2006; Makarova et al., 2011b; Richter et al., 2012a). In contrast to these studies, a genome wide analysis in *E. coli* did not reveal protein–protein interactions between Cas1 and Cas2; however Cas1 interacted with Cas6e and Cas7 (Babu et al., 2011), suggesting that variations likely exist between subtypes.

How are pre-spacers selected and added to arrays?

How do CRISPR-Cas systems selectively acquire foreign DNA from phages and plasmids for incorporation into CRISPR arrays? As mentioned earlier, research in *S. thermophilus* led to the identification of a short sequence adjacent to protospacers in the viral genomes, which was aptly named the protospacer adjacent motif (PAM) (Deveau et al., 2008; Mojica et al., 2009). Subsequently, an extensive bioinformatic approach identified the PAMs for many CRISPR-Cas systems (Mojica et al., 2009). However, the requirement for PAMs in interference had precluded direct independent testing of their role in acquisition. In theory, phage challenge could result in random spacer acquisition, but yet only those that provide productive interference by targeting a viral protospacer with the correct PAM would survive the selection and be detected. Elegant experiments of spacer acquisition in the absence of interference by Yosef et al. (2012) refuted this theory and demonstrated that in *E. coli*, spacer selection requires a PAM. In *E. coli*, the PAM is three nt and in other CRISPR-Cas types PAMs are also only a few nt (Mojica et al., 2009). The acquisition of chromosomally-derived spacers was rare relative to integration from the plasmid, even though a higher number of possible PAM targets existed in the chromosomal DNA (Yosef et al., 2012). Therefore, despite a clear requirement of the PAM for integration, these data suggest that additional discrimination mechanisms must exist (Yosef et al., 2012). What this mechanism(s) entails is a mystery, but it is possible that the presence or absence of some sort of DNA modification is detected. Alternatively, incorporation of chromosomally-derived spacers might be disfavored due to topological constraints between CRISPR arrays, Cas proteins and chromosomal pre-spacers during formation of an integration complex.

Once the spacer has been cut from the foreign DNA, presumably by either, or both, Cas1 and Cas2, what could be the

remaining CRISPR requirements for spacer acquisition? In the type I-E system, a single 'repeat' and 60 bp of the leader 5′ of the 'repeat' was sufficient for incorporation of a new repeat-spacer unit when Cas1 and Cas2 were expressed (Yosef et al., 2012). The involvement of 60 bp of the leader suggests that this region contain sequences recognized by Cas1 and/or Cas2 (and possibly other host-derived factors) that are necessary for incorporation. In addition, it was observed that when the spacer is added to the expanding CRISPR array it delivers the last nt of the PAM, which becomes the most 3′ nt of the leader proximal repeat (Fig. 2A) (Datsenko et al., 2012; Goren et al., 2012; Swarts et al., 2012). This suggests that the nuclease(s) responsible (possibly Cas1 and/or Cas2) cleave the pre-spacer between the -1 and -2 nt of the PAM. The term 'duplicon' has been coined to explain this phenomenon, whereby only the initial 28 nt of the repeat are copied (Goren et al., 2012). It is tempting to speculate that the requirement for the leader and the PAM in incorporation provides the selectivity to enable incorporation of new spacers at the leader proximal end of the CRISPR array and in the correct orientation. However, the duplicon model whereby not the entire sequence of the repeat is copied, may not be a universal model applicable to all CRISPR-Cas systems. Erdmann and Garrett (2012) observed that the final nt of certain PAMs, such as the CCN in *S. solfataricus*, was not conserved and that spacers could be integrated in the wrong orientation. Lopez-Sanchez et al. (2012) also recently described the same observations in *Streptococcus agalactiae*.

Interestingly, transcription is not thought to be important for spacer acquisition by the type I-E system since the 60 bp of leader does not include the CRISPR promoter, yet still facilitates adaptation (Yosef et al., 2012). Despite a single 'repeat' being sufficient for acquisition, this did not demonstrate which repeat is copied in a typical array containing multiple repeats. By analyzing two repeats with a single nt difference, it was demonstrated that the leader proximal repeat is duplicated when new spacers are inserted (Yosef et al., 2012). Although Cas1 and Cas2 are shared among CRISPR-Cas systems, differences in repeats and other protein components will likely result in variations on a theme for spacer incorporation.

How do these systems kill viral escapees?

As explained earlier, spacer acquisition occurs against a virus not previously encountered; this can be considered naïve adaptation (Fig. 2A). However, from the work in *E. coli*, it is evident that adaptation can involve two steps; naïve- and priming-integration (*i.e.* the positive feedback loop) (Datsenko et al., 2012; Swarts et al., 2012). In the process of priming (Fig. 2B), the presence of the first spacer, with complementarity to the viral genome, enables the rapid acquisition of multiple spacers targeting that invader. Interestingly, there is an enrichment of spacers from the same DNA strand as the original spacer, suggesting that the targeting or binding promoted by this spacer is important for the generation of new incorporation substrates (Datsenko et al., 2012; Swarts et al., 2012). By assessing priming in mutant strains, Datsenko et al. (2012) demonstrated that the entire Cas system is required (Cas1, Cas2, Cascade-crRNA and Cas3), but surprisingly, strains with spacers against viruses that have evaded targeting, due to single mutations in the protospacer or PAM, still displayed priming. What would be the advantage of this feedback system? Viruses or plasmids that have acquired point mutations to evade CRISPR-Cas targeting would quickly be detected and, through the integration of new spacers, these elements would be eliminated. Therefore, in the phage-bacterium arms race, CRISPR-Cas defense is less vulnerable to evasion by single point mutations than previously thought. Furthermore, the accumulation of multiple spacers against a single invader can strengthen resistance and reduce the probability of escape, as multiple mutations in the mobile element would be required.

Concluding remarks

Memorization of previous infections is a key characteristic of the CRISPR-Cas system that protects bacteria and archaea against foreign mobile genetic elements. Following infection, a short sequence of invading DNA is incorporated into the CRISPR array of the host genome, generating an ordered memory bank that directs the recognition of the same invaders upon repeated infection. Although the exact molecular mechanisms involved in adaptation are yet to be determined, recent studies have shown that both Cas1 and Cas2 are essential for acquisition and highlighted the critical requirement for the PAM sequence of the candidate pre-spacer (invading DNA) and the CRISPR leader sequence for integration of spacers. A number of questions remain to be addressed: Do Cas1 and Cas2 form a complex? How do they interact with the invader and what is the mechanism to integrate the spacer into the CRISPR array? Are protein(s) other than Cas1 and Cas2 involved in the naïve process of spacer acquisition? Are Cas1, Cas2, Cas3 and the Cascade-crRNA complex the only components required for priming acquisition? Which conditions of infection are a trigger for priming? Are features other than the PAM involved in acquiring foreign DNA rather than self-DNA? The recent development of genetic strategies for studying the process of adaptation promise to reveal answers to many of these questions.

Acknowledgments

CRISPR-Cas research in the Fineran laboratory is supported by a University of Otago Research Grant, the Deans Bequest Fund from the Otago School of Medical Sciences and a Rutherford Discovery Fellowship to PCF from the Royal Society of New Zealand. Research on CRISPR-Cas in the Charpentier laboratory is funded by the Kempe Foundation (#SMK-1136.1), the Swedish Research Council (#K2010-57X-21436-01-3 and #621-2011-5752-LiMS) and Umeå University (#Dnr: 223-2728-10, #Dnr: 223-2836-10 and #Dnr: 223-2989-10). We thank Stan Brouns for his critical comments on the manuscript.

References

Andersson, A.F., Banfield, J.F., 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320, 1047–1050.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E.V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., Yakunin, A.F., 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484–502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712.

Barrangou, R., Horvath, P., 2012. CRISPR: new horizons in phage resistance and strain identification. Annu. Rev. Food Sci. Technol. 3, 143–162.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A., Yakunin, A.F., 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361–20371.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., Yakunin, A.F., 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. EMBO J. 30, 4616–4627.

Bhaya, D., Davison, M., Barrangou, R., 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu. Rev. Genet. 45, 273–297.

Bikard, D., Marraffini, L.A., 2012. Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. Curr. Opin. Immunol. 24, 15–20.
Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561.
Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964.
Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., O'Toole, G.A., 2012. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. J. Bacteriol. August 10 (Epub ahead of print).
Carte, J., Wang, R., Li, H., Terns, R.M., Terns, M.P., 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 22, 3489–3496.
Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945.
Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.
Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390–1400.
Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage–bacteria interactions. Annu. Rev. Microbiol. 64, 475–493.
Drevet, C., Pourcel, C., 2012. How to identify CRISPRs in sequencing data. Methods Mol. Biol. 905, 15–27.
Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S., Kuramitsu, S., 2006. Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. Protein Sci. 15, 1494–1499.
Erdmann, S., Garrett, R.A., 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol. Microbiol. 85, 1044–1056.
Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., Moineau, S., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.
Goren, M.G., Yosef, I., Auster, O., Qimron, U., 2012. Experimental definition of a clustered regularly interspaced short palindromic duplicon in *Escherichia coli*. J. Mol. Biol. 423, 14–16.
Grissa, I., Bouchon, P., Pourcel, C., Vergnaud, G., 2008. On-line resources for bacterial micro-evolution studies using MLVA or CRISPR typing. Biochimie 90, 660–668.
Grissa, I., Vergnaud, G., Pourcel, C., 2007a. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinf. 8, 172.
Grissa, I., Vergnaud, G., Pourcel, C., 2007b. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35, W52–W57.
Groenen, P.M., Bunschoten, A.E., van Soolingen, D., van Embden, J.D., 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol. Microbiol. 10, 1057–1065.
Haft, D.H., Selengut, J., Mongodin, E.F., Nelson, K.E., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. 1, e60.
Han, D., Lehmann, K., Krauss, G., 2009. SSO1450—a CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA. FEBS Lett. 583, 1928–1932.
Hatoum-Aslan, A., Maniv, I., Marraffini, L.A., 2011. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc. Natl. Acad. Sci. U S A 108, 21218–21222.
Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., Doudna, J.A., 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 329, 1355–1358.
Hendrix, R.W., 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6, 506–511.
Horvath, P., Barrangou, R., 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167–170.
Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., Barrangou, R., 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int. J. Food Microbiol. 131, 62–70.
Horvath, P., Romero, D.A., Côuté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401–1412.
Howard, J.A., Delmas, S., Ivančić-Baće, I., Bolt, E.L., 2011. Helicase dissociation and annealing of RNA–DNA hybrids by *Escherichia coli* Cas3 protein. Biochem. J. 439, 85–95.
Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. 169, 5429–5433.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.
Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M.R., Barendregt, A., Zhou, K., Snijders, A.P., Dickman, M.J., Doudna, J.A., Boekema, E.J., Heck, A.J., van der Oost, J., Brouns, S.J., 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. 18, 529–536.
Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317–327.
Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copie, V., Young, M.J., White, M.F., Lawrence, C.M., 2011. Structural and functional characterization of an archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Anti-viral Defense (CASCADE). J. Biol. Chem. 286, 21643–21656.
Lopez-Sanchez, M.J., Sauvage, E., Da Cunha, V., Clermont, D., Ratsima Hariniaina, E., Gonzalez-Zorn, B., Poyart, C., Rosinski-Chupin, I., Glaser, P., 2012. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. Mol. Microbiol. 85, 1057–1071.
Makarova, K.S., Aravind, L., Wolf, Y.I., Koonin, E.V., 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct 6, 38.
Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7.
Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011b. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477.
Marraffini, L.A., Sontheimer, E.J., 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. 11, 181–190.
Masepohl, B., Gorlitz, K., Bohme, H., 1996. Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. Biochim. Biophys. Acta 1307, 26–30.
Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740.
Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182.
Mulepati, S., Bailey, S., 2011. Structural and biochemical analysis of the nuclease domain of the clustered regularly interspaced short palindromic repeat (CRISPR) associated protein 3 (CAS3). J. Biol. Chem. 286, 31896–31903.
Nam, K.H., Ding, F., Haitjema, C., Huang, Q., Delisa, M.P., Ke, A., 2012a. Double-stranded endonuclease activity in *B. halodurans* Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Cas2 protein. J. Biol. Chem. August 31 (Epub ahead of print).
Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., Ke, A., 2012b. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. Structure 20, 1574–1584.
Plagens, A., Tjaden, B., Hagemann, A., Randau, L., Hensel, R., 2012. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. J. Bacteriol. 194, 2491–2500.
Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653–663.
Pride, D.T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White 3rd, R.A., Loomer, P., Armitage, G.C., Relman, D.A., 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J. 6, 915–926.
Pride, D.T., Sun, C.L., Salzman, J., Rao, N., Loomer, P., Armitage, G.C., Banfield, J.F., Relman, D.A., 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. Genome Res. 21, 126–136.
Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B., Fineran, P.C., 2011. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrospecticum*. RNA Biol. 8, 517–528.
Richter, C., Gristwood, T., Clulow, J.S., Fineran, P.C., 2012a. *In vivo* protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas system. PLoS One, in press.
Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., Randau, L., 2012b. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Res. Aug 8. [Epub ahead of print].
Samai, P., Smith, P., Shuman, S., 2010. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66, 1552–1556.
Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res. 39, 9275–9282.
Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P, Siksnys, V., 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. 30, 1335–1342.
Swarts, D.C., Mosterd, C., van Passel, M.W., Brouns, S.J., 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One 7, e35888.
Terns, M.P., Terns, R.M., 2011. CRISPR-based adaptive immune systems. Curr. Opin. Microbiol. 14, 321–327.

Touchon, M., Charpentier, S., Clermont, O., Rocha, E.P., Denamur, E., Branger, C., 2011. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J. Bacteriol. 193, 2460–2467.

Touchon, M., Rocha, E.P., 2010. The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. PLoS One 5, e11126.

Tyson, G.W., Banfield, J.F., 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol. 10, 200–207.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., Brouns, S.J., 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. 34, 401–407.

Vergnaud, G., Li, Y., Gorge, O., Cui, Y., Song, Y., Zhou, D., Grissa, I., Dentovskaya, S.V., Platonov, M.E., Rakin, A., Balakhonov, S.V., Neubauer, H., Pourcel, C., Anisimov, A.P., Yang, R., 2007. Analysis of the three *Yersinia pestis* CRISPR loci provides new tools for phylogenetic studies and possibly for the investigation of ancient DNA. Adv. Exp. Med. Biol. 603, 327–338.

Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28, 127–181.

Westra, E.R., Brouns, S.J., 2012. The rise and fall of CRISPRs—dynamics of spacer acquisition and loss. Mol. Microbiol. 85, 1021–1025.

Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., de Vos, W.M., Dame, R.T., de Vries, R., Brouns, J., van der Oost, J., 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by cascade and Cas3. Mol. Cell 46, 595–605.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., Nogales, E., 2011a. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486–489.

Wiedenheft, B., Sternberg, S.H., Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331–338.

Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., Doudna, J.A., 2011b. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc. Natl. Acad. Sci. U S A 108, 10092–10097.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., Doudna, J.A., 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17, 904–912.

Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40, 5569–5576.